

UNIVERSIDAD ROSARIO  
CASTELLANOS

LICENCIATURA EN CIENCIAS DE DATOS  
PARA NEGOCIOS



---

EVIDENCIA INTEGRADORA

DE LA PUERTA GIRATORIA A LA VÍA DE  
GRADUACIÓN:

*El Caso de Negocio para el Sistema SAREP en la  
Universidad Nacional Rosario Castellanos (UNRC)*

---



---

**EQUIPO 3:**

- DÍAZ HERNÁNDEZ DANIELA
- NÚÑEZ TREJO ANGÉLICA
- PEDROZA CELIS ARIÁN

**GRUPO Y SEDE:**  
PH301-LCDN-UST

# Contenido

## De la Puerta Giratoria a la Vía de Graduación:

<b>El Caso de Negocio para el Sistema SAREP en la Universidad Nacional Rosario Castellanos (UNRC).....</b>	<b>4</b>
<b>1. Resumen Ejecutivo.....</b>	<b>4</b>
1.1. Diagnóstico de la Crisis: Un Desafío de Eficiencia Institucional.....	4
1.2. El Contexto Nacional: Crisis de Deserción y Oportunidad.....	5
1.3. La Solución Basada en Evidencia: SAREP.....	6
1.4. Veredicto de Viabilidad y Recomendación Estratégica.....	7
<b>2. Diagnóstico: La Anatomía de un Problema Sistémico.....</b>	<b>7</b>
<b>2.1. El Costo de la Inacción: Cartografía del Impacto.....</b>	<b>7</b>
<b>2.1.1. El Costo Público.....</b>	<b>7</b>
2.1.2. El Costo Humano.....	8
<b>2.2. Foco Institucional: El Caso Emblemático de la UNRC.....</b>	<b>10</b>
2.2.1. Crecimiento Exponencial vs. Resultados de Egreso.....	10
Las dos Métricas Clave de Eficiencia.....	10
2.2.2. De la Transparencia a la Inteligencia: Una "Brecha de Acción".....	12
2.2.3. El Perfil del Estudiante y la Asimetría de Recursos.....	12
Una Disparidad Estructural en la Carga Docente.....	13
2.2.4 Del Síntoma a la Intervención: La Lógica del Enfoque Predictivo.....	14
El Rendimiento Académico como "Indicador Mediador".....	15
La Mentalidad como "Amortiguador" del Contexto Estructural.....	15
2.2.5. El Fundamento Teórico: La Integración Académica como "Síntoma Observable".....	15
<b>3. Metodología: Un Enfoque Sistemático Basado en Evidencia.....</b>	<b>16</b>
<b>3.1. Fase 1: Extracción Forense y Consolidación de un Dataset Histórico. 16</b>	
<b>3.2. Fase 2: Descubrimiento de Factores de Riesgo y Validación Local.....</b>	<b>17</b>

3.2.1 Descubrimiento de Características mediante Prototipo Predictivo.....	17
3.2.2 Limitaciones.....	19
3.2.3. Diseño de Instrumento de Validación Local.....	19
Diseño del Instrumento.....	20
Método de Muestreo y Tamaño de la Muestra.....	21
Determinación del tamaño de la muestra.....	22
Elección del valor "p" .....	22
Cálculo del tamaño de muestra.....	22
Justificación de la muestra alcanzada.....	23
Limitaciones.....	24
3.2.4. Resultados y Validación Estadística de Hipótesis Locales.....	26
Hipótesis 1: El Rendimiento Académico (El "Síntoma" Principal).....	27
Hipótesis 2: El "Silencio" de los Factores Financieros y la Hipótesis del Mediador.....	28
Hipótesis 3: El Desajuste Vocacional (Estadísticamente Significativo)....	29
Conclusión Metodológica General: Del Síntoma a la Mentalidad.....	30
<b>3.3. Fase 3: Benchmarking de Soluciones para la Identificación de “Palancas de Acción”.....</b>	<b>31</b>
<b>4. Solución Propuesta: El Sistema SAREP.....</b>	<b>31</b>
4.1. Visión y Arquitectura General: El "Waze" del Viaje Académico.....	31
4.2. Diseño de la Base de Conocimiento: Un Modelo de Datos para un Perfil Estudiantil de 360 Grados.....	32
4.4. Componente 2: El Ecosistema de Intervención Proactiva y el Dashboard del Tutor.....	33
Círculo de Intervención Proactiva.....	34
El Dashboard del Tutor: De la Alerta a la Acción.....	34
La Lógica Interna: Un Motor de Reglas Basado en Evidencia.....	35
<b>5. Análisis de Viabilidad y Caso de Negocio.....</b>	<b>36</b>

5.1. Impacto Probado: Evidencia Cuantitativa.....	36
5.2. Análisis de Costos: Construir vs. Comprar.....	<b>36</b>
Tabla 1: Comparativa de Costos y Beneficios (Proyección Anual).....	37
Análisis de TCO (Costo Total de Propiedad):.....	38
<b>6. Recomendaciones Estratégicas y Hoja de Ruta.....</b>	<b>38</b>
6.1. Hoja de Ruta Física para la Implementación.....	38
6.2. Recomendación de Política Institucional.....	39
6.3. Recomendación Táctica: Estabilización e Ingesta Flexible.....	39
<b>7. Conclusión.....</b>	<b>40</b>
<b>8. Referencias.....</b>	<b>41</b>

# **De la Puerta Giratoria a la Vía de Graduación: El Caso de Negocio para el Sistema SAREP en la Universidad Nacional Rosario Castellanos (UNRC)**

## **1. Resumen Ejecutivo**

### **1.1. Diagnóstico de la Crisis: Un Desafío de Eficiencia Institucional**

La Universidad Nacional Rosario Castellanos (UNRC) se erige como un caso emblemático de la tensión entre el acceso masivo a la educación y la necesidad de asegurar la compleción estudiantil. Creada con un mandato para ampliar radicalmente el acceso, la UNRC ha visto un crecimiento explosivo, alcanzando una matrícula de hasta 57,441 estudiantes (UNRC, 2025). Sin embargo, este éxito en la captación contrasta con un profundo desafío de eficiencia.

Esta crisis se manifiesta en un cuello de botella crítico en la titulación: datos oficiales del último trimestre de 2024 revelan que de 5,572 estudiantes que completaron sus créditos (egresados), solo 244 obtuvieron su título, lo que representa una tasa de conversión a titulación anual de apenas el ~4.4% (UNRC, 2024). Este bloqueo en la "última milla" del viaje estudiantil, sumado a una tasa de eficiencia terminal (~5.0%) significativamente inferior a la de sus pares.

Esta cifra es notablemente inferior a la de sus pares nacionales, como la Universidad Autónoma Metropolitana (UAM) con ~8.0% y, de manera más crítica, el Instituto Politécnico Nacional (IPN), que ostenta una tasa del 16.7% (UAM, 2024; IPN, 2024). Esta brecha de más de 10 puntos porcentuales con el IPN no es solo una estadística; representa una pérdida significativa de potencial humano y de retorno sobre la inversión pública.

El problema subyacente no es la falta de datos, sino una brecha en el análisis y la acción. La propia UNRC reporta datos de "bajas definitivas", pero carece de los sistemas para transformar esta información en inteligencia procesable e intervenciones

proactivas (UNRC, 2023). Este desequilibrio pone en grave riesgo la sostenibilidad de la institución y el cumplimiento de su misión fundamental.

#### El Diagnóstico de la Crisis en Cifras (Infografía)<sup>1</sup>



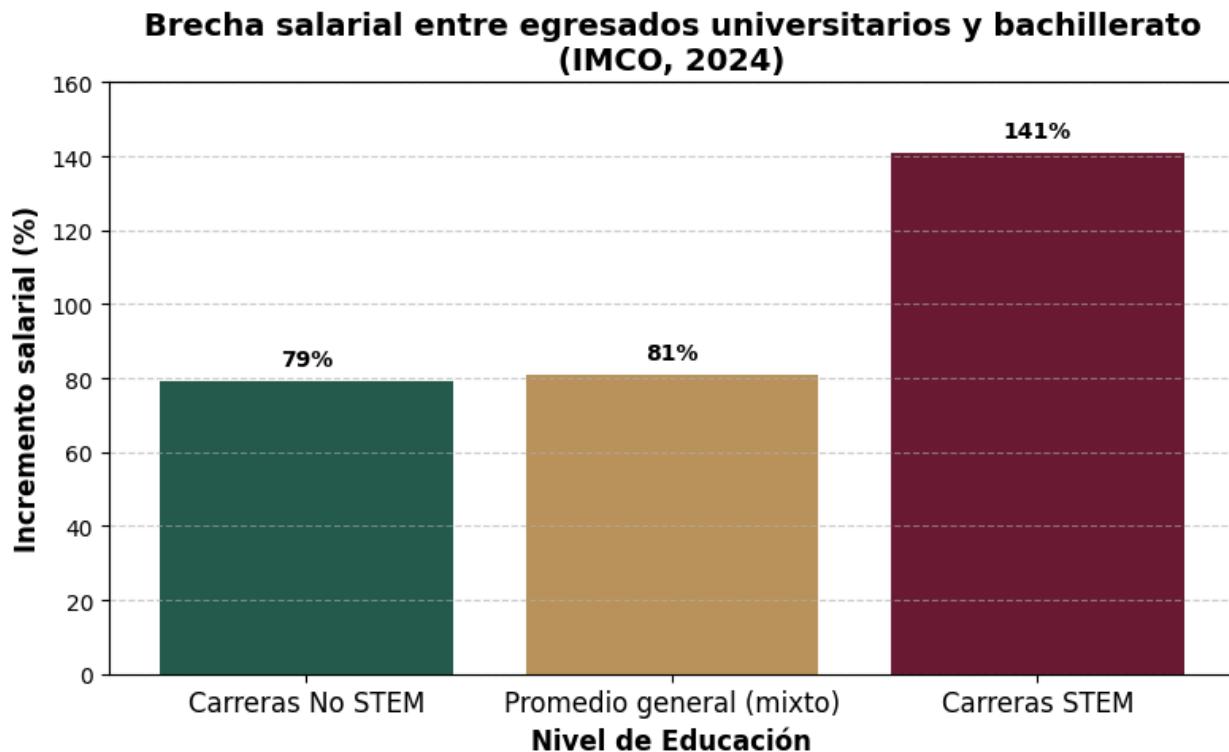
## 1.2. El Contexto Nacional: Crisis de Deserción y Oportunidad

A nivel macroeconómico, el desafío de la UNRC se enmarca en una crisis nacional de deserción en la educación superior. Cifras oficiales del INEGI indican que aproximadamente 307,000 estudiantes abandonan sus estudios universitarios cada año, lo que representa una pérdida de capital humano y fiscal estimada en \$21.3 mil millones de pesos anuales (INEGI, 2024; Castillero et al., 2019).

Más allá del impacto fiscal, la deserción impone un costo devastador a nivel individual y familiar. Representa una pérdida patrimonial multimillonaria derivada de la brecha salarial del 81% que existe entre titulados y no titulados (IMCO, 2024). Este diferencial alcanza su máxima expresión en áreas estratégicas como matemáticas, economía y ciencias, donde los ingresos son hasta 141% superiores al promedio nacional (IMCO, 2024), consolidando un retorno económico y social que trasciende al individuo y fortalece la competitividad del país.

<sup>1</sup> Fuente: Elaboración propia con base en datos de UNRC (Numeralia, 2021-2025), UAM, 2024; IPN, 2024.

Gráfico de barras comparativas con los porcentajes de brecha salarial



Fuente: Elaboración propia con base en datos del Instituto Mexicano para la Competitividad, A.C. (IMCO, 2024)

En una sociedad como la mexicana, la educación es el principal motor de movilidad social, y la deserción es el principal mecanismo de fractura de esta promesa de ascenso social, perpetuando la desigualdad (CEEY, s.f.).

### 1.3. La Solución Basada en Evidencia: SAREP

Este informe presenta el caso de negocio para la implementación SAREP, el Sistema de Acompañamiento y Retención Estudiantil Proactivo. SAREP no es un mero software, sino un ecosistema estratégico diseñado para:

- **Predecir** el riesgo de abandono de cada estudiante mediante un modelo de Machine Learning validado.
- **Intervenir** de forma temprana a través de un ecosistema coordinado de apoyos, activado por alertas predictivas.

- **Demostrar** su valor a través de un caso de negocio que cuantifica el retorno sobre la inversión pública.

## 1.4. Veredicto de Viabilidad y Recomendación Estratégica

La implementación de SAREP es viable, esencial y estratégicamente superior. La evidencia internacional de instituciones como Georgia State University (GSU), que aumentó su tasa de graduación en 23 puntos porcentuales, y de plataformas como EAB Navigate360, con un ROI probado de 5:1, demuestra que estos sistemas funcionan (GSU, s.f.; EAB, s.f.).

La recomendación es inequívoca: la UNRC debe adoptar SAREP para transitar de una “política de acceso” a una “economía de la compleción”, asegurando su misión y la viabilidad de la inversión social.

## 2. Diagnóstico: La Anatomía de un Problema Sistémico

### 2.1. El Costo de la Inacción: Cartografía del Impacto

#### 2.1.1. *El Costo Público*

La inversión del Estado en la educación superior es una apuesta por el capital humano. Cada estudiante que deserta convierte un activo potencial en una pérdida neta. El costo económico promedio no recuperable se ha calculado en \$69,411 MXN por estudiante (Castillero et al., 2019). Extrapolando esta cifra a los aproximadamente 307,000 desertores anuales, la pérdida económica directa para el país asciende a \$21.3 mil millones de pesos (INEGI, 2024).

Esta cifra, sin embargo, subestima el impacto. La deserción distorsiona métricas clave como el “costo por egresado”. Si una universidad recibe presupuesto para 1,000 estudiantes, pero solo gradúa a 500, el costo real se duplica. Este indicador es monitoreado por la SEP y la SHCP, y puede justificar recortes, creando un ciclo vicioso de bajo rendimiento y menor financiamiento.

### **2.1.2. El Costo Humano**

Más allá del impacto fiscal, la deserción universitaria impone un costo devastador en la esfera individual y familiar. Representa una pérdida patrimonial multimillonaria derivada de la brecha salarial del 81% entre titulados y no titulados, brecha que se amplifica a lo largo de toda una vida laboral (Instituto Mexicano para la Competitividad, 2024). Este diferencial alcanza su máxima expresión en áreas estratégicas como matemáticas, economía y ciencia de datos, donde los ingresos son hasta 141% superiores al promedio nacional, consolidando un retorno económico y social que trasciende al individuo y fortalece la competitividad del país.

Sin embargo, el costo más profundo no es financiero, sino social. En una sociedad como la mexicana, la educación superior es el principal motor de la movilidad social, donde 74 de cada 100 personas que nacen en la pobreza no logran superarla (CEEY, s.f.), la educación superior es el principal motor de movilidad social. La deserción, por lo tanto, no es solo un evento académico; es el principal mecanismo de fractura de esta promesa de ascenso social, perpetuando la desigualdad intergeneracional y truncando cientos de miles de trayectorias de vida cada año (CEEY, s.f.).

El impacto más profundo de la deserción se manifiesta en la exposición del individuo a un mayor riesgo de informalidad laboral. Según datos del INEGI y Data México, el mercado laboral en México está fuertemente segmentado por nivel educativo (Secretaría de Economía, s.f.).

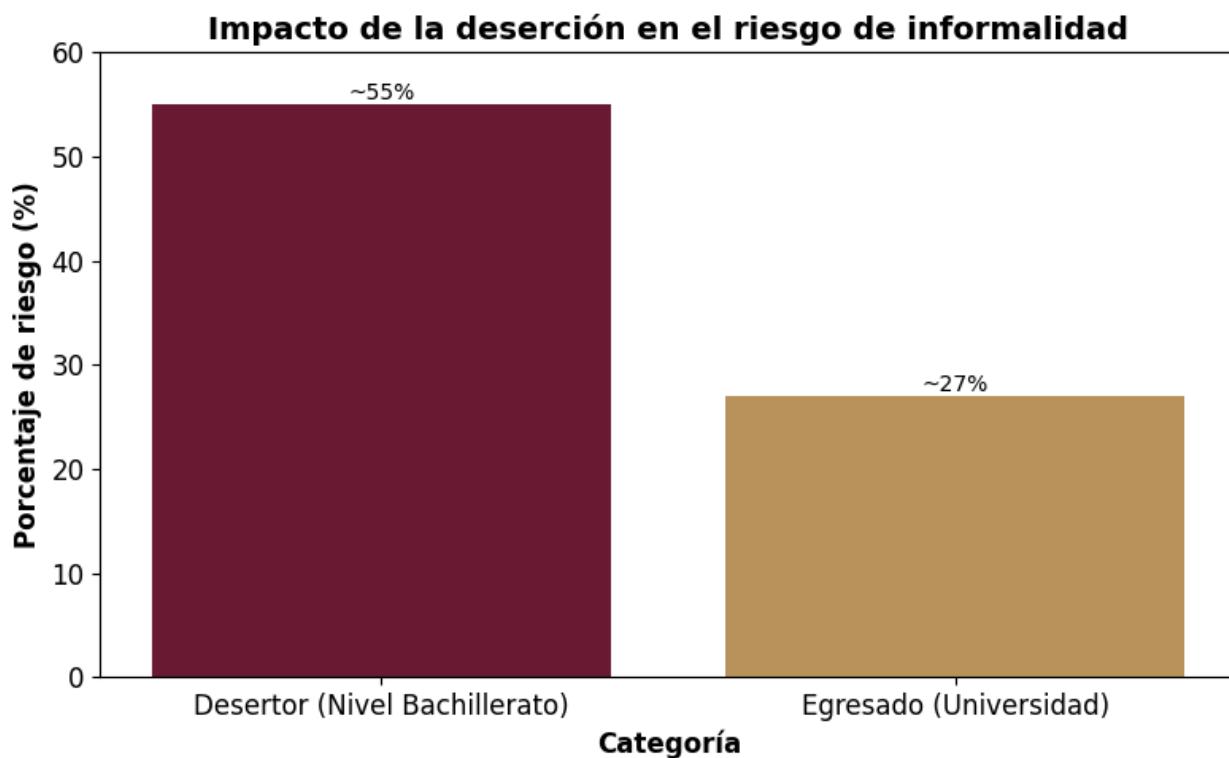
Un individuo con licenciatura completa ingresa al grupo de "Profesionistas y Técnicos", donde la tasa de informalidad es de solo un 26.8%.

Un desertor universitario, al no poseer un título, compite en el mercado laboral cuyo máximo grado es la educación media superior, un segmento donde la tasa de informalidad es análoga al promedio nacional, cercana al 54.8%.

La informalidad implica la exclusión automática de la seguridad social, las pensiones y el crédito hipotecario, herramientas fundamentales para la construcción de un

patrimonio intergeneracional. Por lo tanto, la deserción no es solo un evento académico; es una sentencia que puede perpetuar el ciclo de la desventaja.

**Exhibit 2: Mayor Riesgo de Informalidad**



Fuente: Elaboración propia

## **2.2. Foco Institucional: El Caso Emblemático de la UNRC**

### ***2.2.1. Crecimiento Exponencial vs. Resultados de Egreso***

La UNRC ha cumplido con creces su mandato de ampliar el acceso. Su matrícula ha experimentado un crecimiento explosivo, pasando de 25,429 a inicios de 2021 a 57,441 para el segundo trimestre de 2025.

Es decir, la universidad ha alcanzado una matrícula que ha crecido más de un 121% desde 2021. Sin embargo, este éxito en la captación enmascara un profundo desafío de eficiencia terminal. Un análisis de los datos históricos revela una dinámica de 'puerta giratoria': desde 2022 hasta 2025, la institución ha registrado 18,636 bajas definitivas, en contraste con solo 970 titulados. Esto se traduce en una alarmante ratio de casi 20

estudiantes que abandonan por cada uno que se titula, y una tasa de finalización histórica efectiva de apenas el 5%.

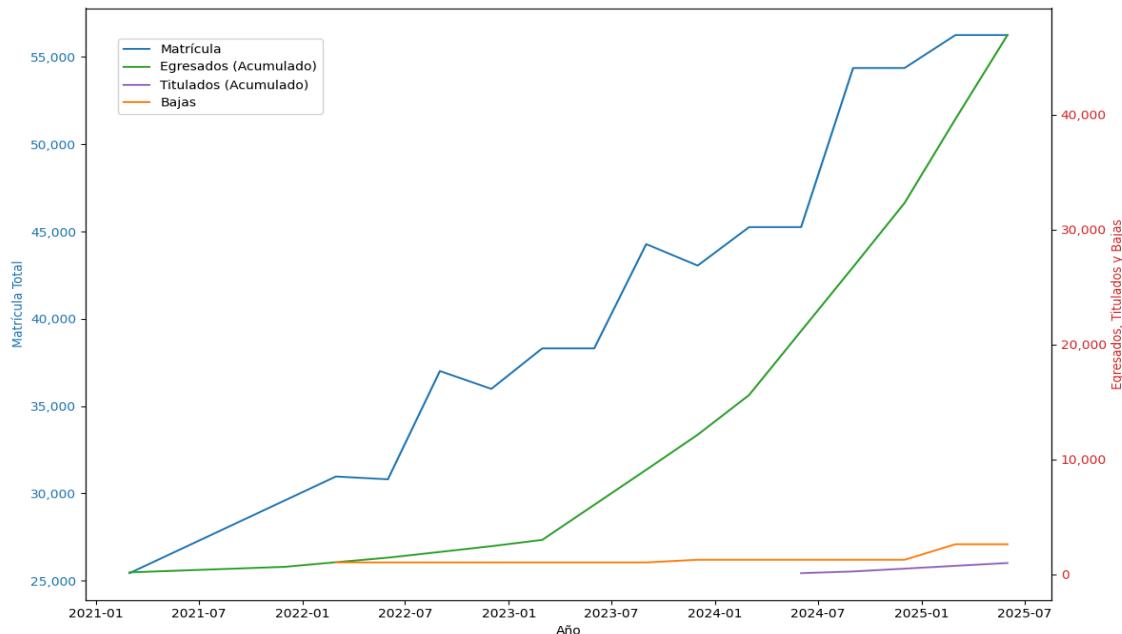
## Las dos Métricas Clave de Eficiencia

Métrica	Cálculo	Nos dice	Insight Clave
<b>Tasa de Egreso Anual (2024 Q4)</b>	(Matriculados Totales en 2024) / (Egresados Totales)	¿Qué tan eficiente es el proceso administrativo final?	4.38% (Extremadamente ineficiente, un "cuello de botella") comparada con pares como el IPN al 16.7.
<b>Tasa de Finalización Efectiva</b>	(Titulados Totales) / (Titulados + Bajas)	¿Cuál es la probabilidad histórica de éxito vs. fracaso?	~5% (El resultado "Baja" es 19 veces más probable que "Titulado")

Sin embargo, este éxito en la captación contrasta dramáticamente con sus resultados de egreso. El desafío se centra en el cuello de botella de la titulación. Datos del último informe 'Numeralia' (Q4 2024) revelan que de 5,572 egresados, solo 244 han logrado obtener su título, una tasa de conversión a titulación de apenas el ~4.4%. Este bloqueo en la 'última milla' es el principal desafío de rendimiento de la institución.

El siguiente gráfico, de tendencia histórica longitudinal, visualiza esta dinámica, contando la historia completa de la "puerta giratoria" institucional en una sola imagen.

**Evolución Histórica de Métricas Clave - 'Crecimiento Desacoplado' (2021-2024, Datos Trimestrales)**



Fuente: Elaboración propia con base en datos de UNRC (Numeralia,2021-2025)

El análisis visual del gráfico es revelador. Mientras que la línea de Matrícula Total (azul) muestra una pendiente ascendente y volátil, reflejando una captación masiva, las líneas de resultados cuentan una historia muy diferente. La línea de Egresados Acumulados (verde) crece de manera constante, indicando que los estudiantes están completando sus créditos.

Sin embargo, el hallazgo más crítico se encuentra en la línea de Titulados Acumulados (morada), que permanece casi plana. Esta inmensa brecha visual entre los egresados y los titulados es la evidencia irrefutable de un severo cuello de botella en la titulación. Confirma los datos puntuales que revelan una tasa de conversión a titulación de apenas el 4.4%. Este bloqueo en la fase del proceso estudiantil —el proceso administrativo para obtener el título— es, en la actualidad, el principal desafío de eficiencia y rendimiento de la institución. La línea de Bajas (naranja), aunque de menor magnitud, muestra una acumulación constante de estudiantes que abandonan, contribuyendo al diferencial entre la matrícula y los resultados finales.

### ***2.2.2. De la Transparencia a la Inteligencia: Una "Brecha de Acción"***

La investigación revela que el principal obstáculo para la gestión estratégica en la UNRC no es una ausencia total de métricas, sino una "brecha de acción": la diferencia entre tener datos y transformarlos en inteligencia proactiva.

La propia existencia del portal 'Numeralia' demuestra la intención de la institución de ser transparente, recopilando y publicando periódicamente indicadores clave como matrícula, egresados y bajas definitivas. Sin embargo, estos datos se encuentran dispersos en informes trimestrales, sin un repositorio histórico centralizado ni un análisis público que permita visualizar tendencias.

Esta situación pone de manifiesto una tensión sistémica clave: la tensión entre la "política del acceso" y la "economía de la compleción". La comunicación institucional prioriza la métrica del acceso (nuevos estudiantes), mientras que los indicadores de éxito (retención, graduación) permanecen como datos brutos, sin análisis que guíen la

estrategia. Es importante subrayar que la no publicación de datos a nivel individual no es una falla, sino un cumplimiento de la ley de protección de datos. La brecha, por tanto, no está en la transparencia, sino en la ausencia de un sistema robusto que, respetando la privacidad, aproveche estos datos agregados para generar inteligencia y orientar las intervenciones. SAREP está diseñado precisamente para llenar ese vacío.

### ***2.2.3. El Perfil del Estudiante y la Asimetría de Recursos***

El diseño de una solución efectiva exige comprender no solo el "qué" del problema, sino también el "quién" y el "porqué". El perfil del estudiante de la UNRC y la asignación de recursos institucionales revelan una dinámica que justifica plenamente la necesidad de un sistema proactivo y tecnológico como SAREP.

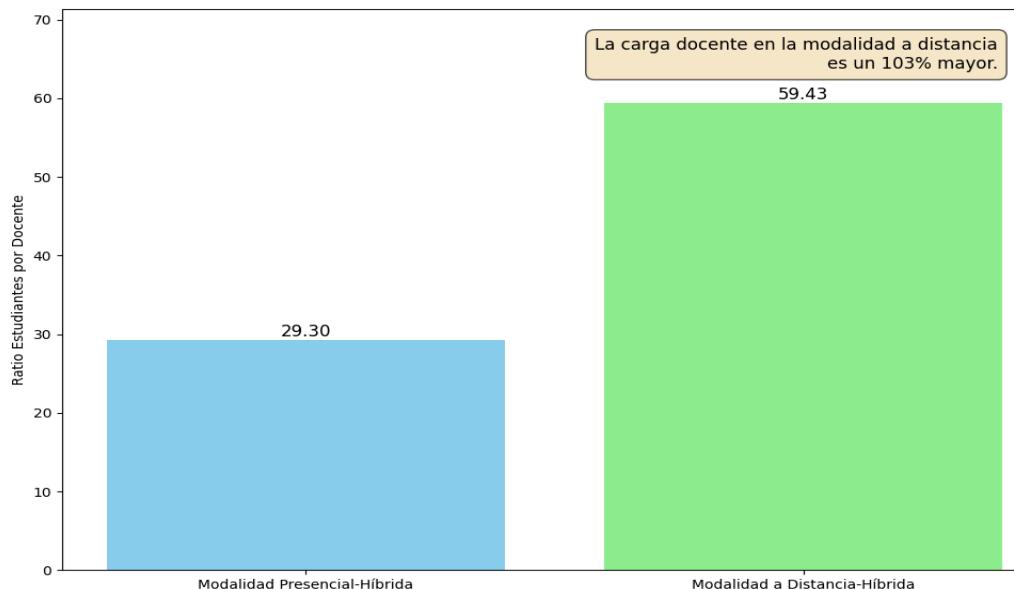
La UNRC atiende predominantemente a un gran número de "estudiantes no tradicionales" (Rodríguez, 2022), quienes a menudo deben equilibrar sus estudios con responsabilidades laborales y familiares. El perfil demográfico revela una notable preponderancia de mujeres (63%) y, de manera crucial, un predominio de la modalidad a distancia o híbrida, que concentra a más del 60% del alumnado. Si bien esta flexibilidad es un pilar del éxito en el acceso, introduce desafíos significativos de aislamiento y autogestión. Este perfil, por su naturaleza, exige un sistema de apoyo digital y proactivo, ya que los modelos de acompañamiento tradicionales resultan insuficientes.

### ***Una Disparidad Estructural en la Carga Docente***

Este desafío se ve exacerbado por una disparidad estructural y persistente en la asignación de recursos. Para evaluar esta dinámica, se realizó un análisis de la ratio estudiante-docente por modalidad. Aunque los informes trimestrales no siempre desglosan la matrícula por modalidad, el Informe de Resultados 2019-2024 de la propia universidad confirma que la modalidad a distancia-híbrida es la predominante, albergando a más del 60% del alumnado a finales de 2023 y 2024 (UNRC, 2024).

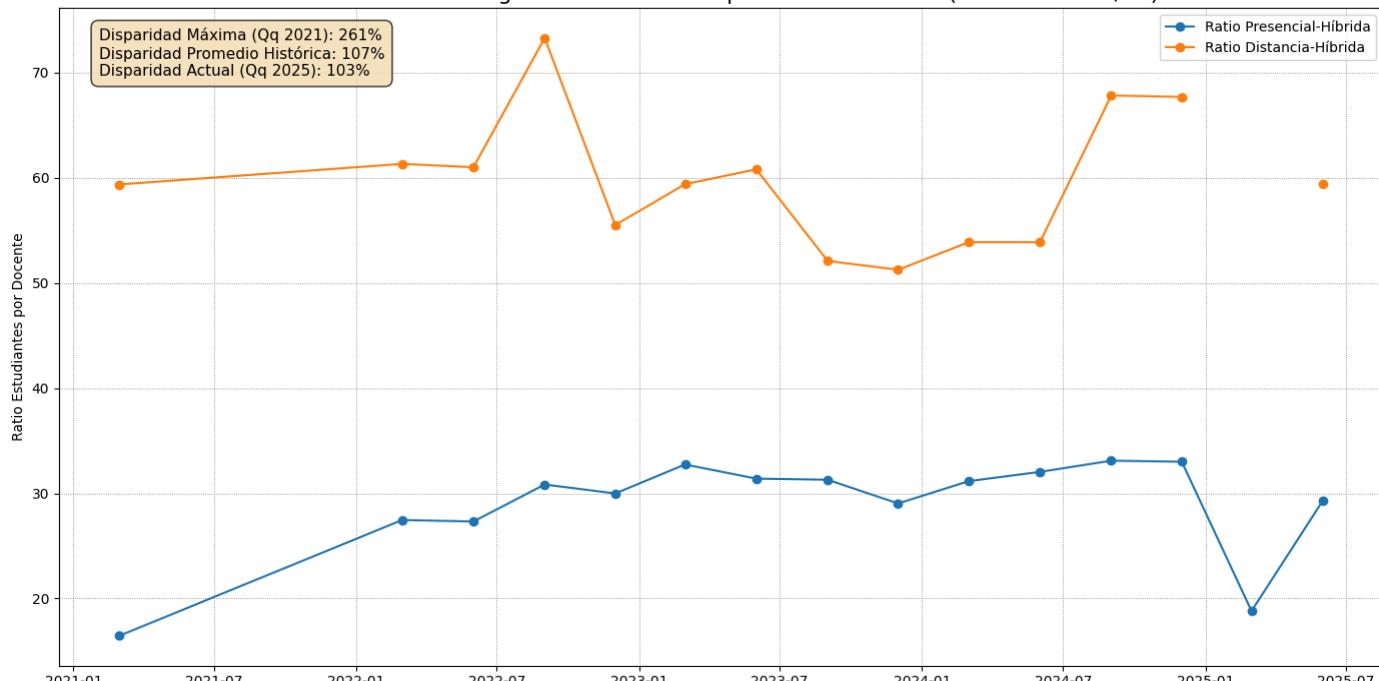
Los resultados, visualizados en el siguiente gráfico, revelan una disparidad significativa y persistente.

### Disparidad en la Carga Docente por Modalidad estimación 60/40 (Q4 2024)



Manteniendo esta proporción histórica constante para los cálculos de 2025, el análisis longitudinal revela que la carga docente en la modalidad a distancia ha sido consistentemente superior a la de la modalidad presencial, alcanzando picos de disparidad de hasta un 107% en períodos anteriores. En el trimestre más reciente (Q1 2025), la ratio en la modalidad a distancia se estima en 59.4 estudiantes por docente, una cifra que es un 103% mayor que la de la modalidad presencial (29.3).

### Evolución de la Carga Docente: Una Disparidad Estructural (Estimación 60/40)



Esta sobrecarga crónica limita severamente la capacidad del personal docente para ofrecer un acompañamiento personalizado y proactivo, precisamente al segmento más grande y de mayor riesgo de la población estudiantil. Este hallazgo constituye una de las justificaciones operativas más fuertes para la implementación de una herramienta tecnológica como SAREP, diseñada para permitir un seguimiento efectivo y a escala.

#### ***2.2.4 Del Síntoma a la Intervención: La Lógica del Enfoque Predictivo***

Si bien el diagnóstico se centra en indicadores de rendimiento, podría argumentarse que factores psicosociales como la motivación o estructurales como el ingreso familiar son las verdaderas causas raíz. Sin embargo, la evidencia científica revela una dinámica más compleja y matizada, que justifica plenamente el enfoque de SAREP en el monitoreo del rendimiento académico temprano.

#### **El Rendimiento Académico como "Indicador Mediador"**

El influyente meta-análisis de Robbins et al. (2004) demostró que el impacto de los factores psicosociales (como la motivación y la autoeficacia) sobre la persistencia está fuertemente mediado por el rendimiento académico real. Es decir, la motivación de un estudiante se traduce en éxito principalmente a través de la obtención de buenas calificaciones. Esto posiciona al rendimiento académico como el punto de convergencia donde la psicología y el comportamiento se encuentran con los resultados observables y medibles. Por lo tanto, un sistema que detecta una caída en las calificaciones está, de hecho, detectando una "fiebre": una señal de alerta temprana de que los factores psicosociales o académicos subyacentes pueden estar fallando. SAREP no intenta ser un psicólogo; actúa como un sistema de diagnóstico temprano que identifica este síntoma crítico para que una intervención humana determine la causa.

#### **La Mentalidad como "Amortiguador" del Contexto Estructural**

De manera similar, la primacía del rendimiento no invalida la importancia de los factores estructurales. Investigaciones de vanguardia, como investigadores de Stanford corroboraron, a nivel nacional, Claro, Paunesku y Dweck (2016), han demostrado que

factores psicológicos como la "mentalidad de crecimiento" —la creencia de que la inteligencia puede desarrollarse— pueden ser predictores del éxito académico tan potentes como el ingreso familiar. De manera crucial, su estudio encontró que la mentalidad de crecimiento actúa como un "amortiguador" contra los efectos de la pobreza, permitiendo a estudiantes de bajos ingresos alcanzar niveles de rendimiento comparables a los de sus pares mucho más ricos.

### ***2.2.5. El Fundamento Teórico: La Integración Académica como "Síntoma Observable"***

Para comprender la raíz de este desafío, es crucial recurrir al marco teórico más influyente en este campo: el Modelo de Integración Estudiantil (SIM) de Vincent Tinto (Tinto, 1993). La teoría de Tinto postula que la persistencia depende de la integración del estudiante en los dominios social y académico. Si bien ambos son importantes, la integración académica tiene un peso preponderante, manifestándose en resultados concretos como el rendimiento en las calificaciones.

Esta primacía del rendimiento temprano no es solo una teoría, sino una conclusión empírica robusta en el campo de la Minería de Datos Educativos (EDM). Estudios como el de Asif et al. (2017) muestran empíricamente que las métricas del primer año son predictores más fuertes que factores contextuales o estructurales, un hallazgo que ha sido replicado en múltiples estudios de EDM, consolidándose como una conclusión robusta en el campo.

Por lo tanto, cuando nuestro modelo SAREP se enfoca en la "señal de oro" del rendimiento temprano, no está midiendo un evento aislado; está detectando el síntoma observable de un proceso de integración o desintegración, permitiendo una intervención fundamentada tanto en la teoría sociológica como en la evidencia de la ciencia de datos.

### **3. Metodología: Un Enfoque Sistemático Basado en Evidencia**

Este informe se sustenta en un proceso riguroso de adquisición, síntesis y validación de evidencia. La metodología se ejecutó en tres fases principales, diseñadas para diagnosticar el problema a nivel macro y luego validar una solución a nivel micro.

#### **3.1. Fase 1: Extracción Forense y Consolidación de un Dataset Histórico**

El primer paso consistió en construir una visión cuantitativa de la eficiencia de la UNRC a lo largo del tiempo. Para ello, se empleó un marco de investigación iterativo que culminó en una misión de extracción forense de datos. Este proceso transformó los datos agregados y dispersos de los informes trimestrales "Numeralia" (2021-2025) en un único dataset histórico longitudinal a nivel institucional.

Este activo, que contiene la serie de tiempo de los indicadores clave (matrícula total, egresos, titulaciones y bajas), permitió realizar un análisis dinámico de las tendencias y se convirtió en la fuente principal de evidencia para el diagnóstico del "crecimiento desacoplado" y la "brecha de eficiencia" presentados en la Sección 2. Paralelamente, se ejecutó una misión de auditoría ("fact-checking") para validar y corregir todas las cifras.

#### **3.2. Fase 2: Descubrimiento de Factores de Riesgo y Validación Local**

Una vez establecido el problema a nivel macro, la segunda fase se centró en responder: ¿Cuáles son los factores de riesgo a nivel individual que impulsan estas tendencias? Ante la ausencia de datos históricos a nivel de estudiante en la UNRC, se diseñó una metodología de dos etapas:

##### ***3.2.1 Descubrimiento de Características mediante Prototipo Predictivo***

Para de-riesgar el proyecto, se adoptó una práctica estándar en la industria: el uso de un dataset proxy para realizar una Prueba de Concepto (PoC).

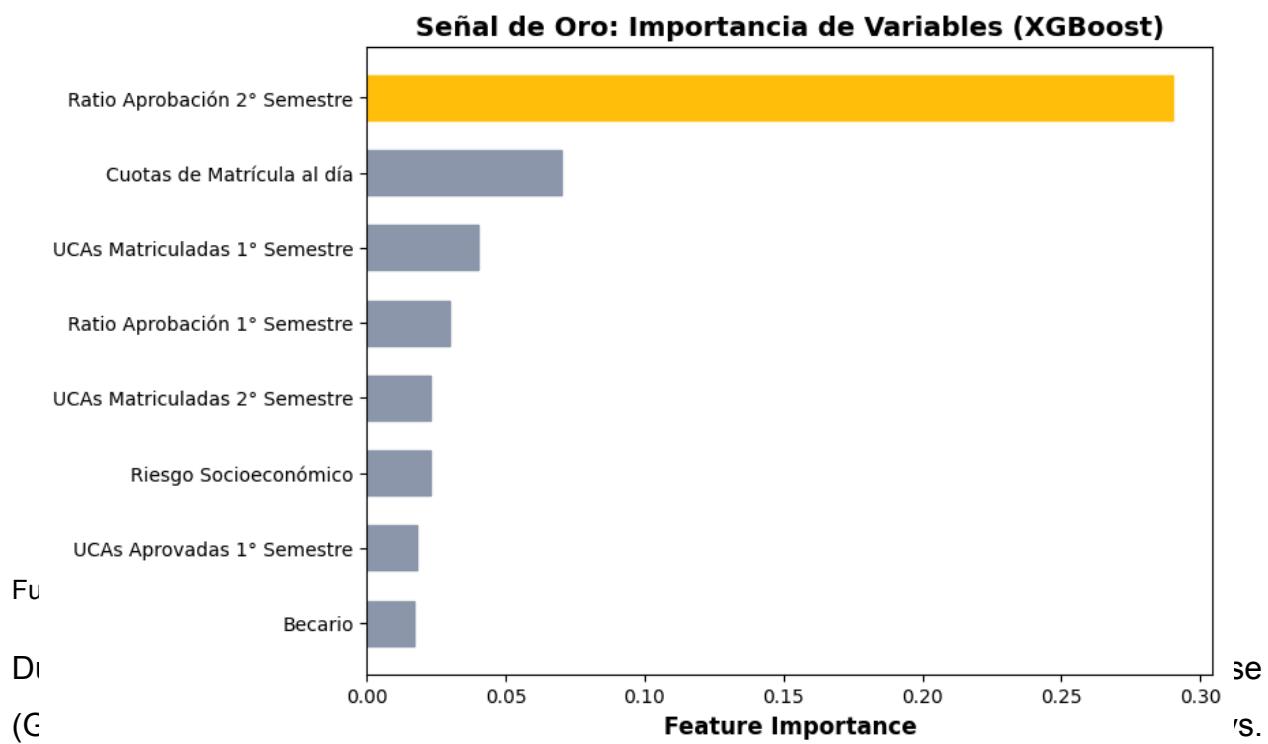
Aplicamos principios de Domain Adaptation. Utilizamos un dataset externo como Dominio Fuente para identificar los patrones latentes universales de la deserción. Luego, validamos la transferibilidad de estos patrones al Dominio Objetivo (UNRC) mediante la encuesta local, confirmando que variables como el rendimiento académico y el desajuste vocacional son predictores transversales, independientemente de la geografía. El propósito no era construir un modelo para la UNRC, sino utilizar un prototipo de modelo XGBoost como una herramienta de descubrimiento.

Se seleccionó el dataset "Predict Students' Dropout and Academic Success" (+4,400 registros) por contener las variables fundamentales que la literatura especializada ha señalado consistentemente como los predictores más robustos de la deserción: el rendimiento académico como medida de la integración académica y el estatus socioeconómico como un factor determinante de la capacidad de integración del estudiante (Tinto, 1993; Robbins et al., 2004).

Sobre este dataset se construyó un prototipo de modelo predictivo utilizando el algoritmo XGBoost. El propósito de este modelo no era generar predicciones para la UNRC, sino actuar como una herramienta de descubrimiento de características. El modelo final alcanzó un Área Bajo la Curva (AUC) promedio de 0.93, validando con alta confianza la viabilidad técnica del enfoque.

El análisis de Feature Importance (ver Gráfico N) reveló un patrón claro, identificando el rendimiento académico en los dos primeros semestres como la "señal de oro" predictiva, muy por encima de otros factores psicosociales estructurales.

**Gráfico N: La "Señal de Oro" - Importancia de Características del Modelo Predictivo**



Retención).

El enfoque multiclase mostró un rendimiento limitado (F1-Score de 0.48 para la clase 'Dropout') debido a la ambigüedad de la categoría 'Matriculado' (Enrolled), que introducía ruido en las fronteras de decisión. Sin embargo, al alinear el modelo con el objetivo de negocio de SAREP —que es la detección binaria de riesgo crítico— se optimizó la arquitectura.

**Resultado Final Optimizada:** Al reestructurar el problema como una clasificación binaria (Target: Dropout), el modelo alcanzó una robustez notable con un F1-Score de 0.81, una Precisión del 85% y una Sensibilidad (Recall) del 78%. Esto confirma que, si bien predecir el estado académico exacto es complejo, detectar la señal de abandono inminente es altamente viable y preciso.

### 3.2.2 Limitaciones

El presente prototipo analítico fue construido utilizando un dataset proxy estandarizado para validar la arquitectura del modelo XGBoost y el flujo de datos. Si bien la validación estadística local (Sección 3.2.3) confirma la correlación de las variables clave en la

población de la UNRC, es condición crítica para la implementación que el modelo sea re-entrenado con el histórico académico real de la institución durante la Fase 1 del Roadmap. Esto transformará las tendencias identificadas en predicciones precisas a nivel individual.

### ***3.2.3. Diseño de Instrumento de Validación Local***

Para la validación local del modelo predictivo SAREP, se diseñó un instrumento denominado Encuesta Pulso Estudiantil UNRC, orientado a captar percepciones, hábitos académicos y factores asociados al riesgo de abandono. Este instrumento se aplicó a una muestra representativa de estudiantes de la Unidad Académica Casco Santo Tomás, sede seleccionada por su carácter mixto (Presencial-Híbrida y a Distancia-Híbrida) y por albergar una alta diversidad disciplinar —en ella se imparten 10 de las 16 licenciaturas ofertadas por la Universidad Nacional Rosario Castellanos (URC, Informe de Resultados 2019–2024, p. 32).

#### **Diseño del Instrumento**

El cuestionario constaba de preguntas clave derivadas del análisis de Feature Importance del modelo XGBoost (entrenado preliminarmente con datos proxy estandarizados), las preguntas sobre satisfacción vocacional derivadas de la revisión de literatura (Tinto, Robbins), planteadas como hipótesis sobre la posible “causa raíz” de ese bajo rendimiento.

El objetivo de este instrumento no fue realizar un censo poblacional, sino ejecutar un sondeo de validación de características (Feature Validation Probe) para confirmar si las variables priorizadas por el algoritmo global (como el desajuste vocacional) mantenían su relevancia estadística en el contexto local.

Su formato breve (menos de 2 minutos) fue diseñado para maximizar la tasa de respuesta y enfocar el análisis en las hipótesis más prometedoras:

**De la Dicotomía a la Magnitud:** Para medir nuestra variable objetivo —la intención de abandono—, no solo nos limitamos a diseñar preguntas dicotómicas (Sí/No) que podría

simplificar en exceso un proceso longitudinal y estar sujeta a sesgos. En su lugar, se utilizó una escala ordinal de 5 puntos para medir la frecuencia e intensidad de los pensamientos de abandono. Este enfoque genera datos más ricos, permitiéndonos analizar no solo la presencia de riesgo, sino también su magnitud.

**Estrategia de Análisis Dual:** Se empleó una doble estrategia de pruebas de hipótesis para cada factor de riesgo:

Una Prueba de Chi-Cuadrado de Independencia para evaluar la asociación con la decisión binaria de considerar el abandono (colapsando la escala ordinal).

Una Prueba No Paramétrica (Kruskal-Wallis o Mann-Whitney U) para analizar el impacto en la intensidad y frecuencia de dichos pensamientos (utilizando la escala ordinal completa).

Este enfoque permitió:

**Maximizar la tasa de respuesta:** gracias a su estructura corta, se obtuvo una muestra efectiva de 100 respuestas válidas.

**Enfocar el análisis:** las preguntas fueron diseñadas para comprobar hipótesis directamente vinculadas con los factores de riesgo más relevantes identificados por el modelo.

- **Conectar el descubrimiento con la validación:** se creó un puente metodológico entre los hallazgos del modelo predictivo y la evidencia empírica local obtenida dentro de la misma institución.

### Método de Muestreo y Tamaño de la Muestra

El método empleado fue muestreo aleatorio simple, donde cada estudiante tuvo la misma probabilidad de ser seleccionado. Se eligió esta técnica por su neutralidad, simplicidad operativa y por garantizar que la muestra representara razonablemente la diversidad de estudiantes de la sede.

La población objetivo corresponde a 1,890 estudiantes, estimación derivada de la capacidad instalada de la Unidad Académica Santo Tomás. Este dato proviene del Informe de Resultados 2019–2024 actualizado, página 30, donde se reporta:

“La Unidad Académica Santo Tomás (...) cuenta con 54 aulas, cada una con capacidad para 35 estudiantes...”

A partir de esta información:

$$N = 54 * 35 = 1,890 \text{ estudiantes}$$

Este valor corresponde a la población máxima operativa de la sede y es adecuado para efectos del cálculo muestral en esta validación local. La elección de esta sede se justifica porque su composición académica refleja la diversidad institucional de la UNRC: agrupa programas de distintas áreas de conocimiento, turnos y modalidades, constituyendo un escenario idóneo para validar localmente las hipótesis derivadas del análisis predictivo.

### Determinación del tamaño de la muestra

Para estimar el tamaño de muestra se utilizaron los siguientes parámetros:

- **Nivel de confianza:** 90%
- **Coeficiente Z:** 1.645
- **Proporción esperada:**  $p = 0.5$
- **Margen de error deseado:**  $E = 0.07 (\pm 7\%)$
- **Población:**  $N = 1,890$

### Eleción del valor “p”

Se empleó la proporción  $p = 0.5$ , conocida como *proporción conservadora*. Este valor representa la máxima variabilidad posible en una variable dicotómica, porque:

$$p(1 - p) = 0.25 \text{ cuando } p = 0.5$$

Esto asegura que el tamaño de muestra sea suficiente aún en el peor escenario (máxima incertidumbre), especialmente cuando no se dispone de datos previos sobre la proporción real del fenómeno en estudio.

### Cálculo del tamaño de muestra

Fórmulas aplicadas:

$$n_0 = \frac{Z^2 \cdot p(1-p)}{E^2}$$

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

#### 1. Cálculo inicial sin corrección por población finita

Sustituyendo valores:

$$n_0 = \frac{(1.645)^2(0.25)}{0.07^2}$$

$$n_0 = \frac{2.706 \cdot 0.25}{0.0049} \approx 138.06$$

#### 2. Cálculo con corrección por población finita

$$n = \frac{138.06}{1 + \frac{137.06}{1890}}$$

$$n = \frac{138.06}{1.0725} \approx 128.7$$

Sustituyendo los valores se obtiene un tamaño de muestra  $n \approx 129$ , que al redondearse hacia arriba y añadir un pequeño colchón operativo para posibles pérdidas de campo o respuestas incompletas, se estableció una meta de 135 respuestas válidas.

### Justificación de la muestra alcanzada

La encuesta obtuvo 100 respuestas válidas, que constituyen una muestra estadísticamente aceptable para el análisis inferencial planificado. Para evaluar la precisión real alcanzada, se calculó el nuevo margen de error con  $n = 100$ , manteniendo los parámetros originales ( $Z = 1.645$ ,  $p = 0.5$ ):

#### Fórmula aplicada:

$$E = Z \cdot \sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Sustituyendo valores:

$$E = 1.645 \cdot \sqrt{\frac{0.25}{100}} \cdot \sqrt{\frac{1890 - 100}{1889}}$$

$$E \approx 1.645 \cdot 0.05 \cdot 0.9734 = 0.0800$$

Aunque el cálculo teórico inicial sugería  $n \sim 129$ , la obtención de 100 respuestas válidas ajusta el margen de error a un +8.0%. Este valor se mantiene dentro de los rangos aceptables para estudios de validación de prototipos y análisis de asociación (no

inferencia poblacional exacta), permitiendo detectar tendencias significativas con robustez.

Por lo tanto, lejos de invalidar la metodología, la obtención de 100 respuestas se reconoce como:

1. **Un tamaño muestral suficientemente amplio** para sostener inferencias con un nivel de confianza del 90% y un margen de error de  $\pm 8\%$ .
2. **Una muestra compatible** con los estándares de estudios educativos aplicados, donde se privilegia la interpretación de patrones y asociaciones locales.
3. **Una decisión metodológica coherente** con el propósito validación de características del estudio, en el que se busca validar hipótesis derivadas de un modelo predictivo más que estimar parámetros poblacionales exactos.

### Limitaciones

A pesar de la rigurosidad del diseño, se reconocen las siguientes limitaciones metodológicas:

1. **Ausencia de Estratificación:** Debido al diseño aleatorio simple y la protección de anonimato, la muestra no garantiza una proporcionalidad exacta por carrera o semestre. Sin embargo, esto no compromete la validez de las pruebas de independencia (Chi-cuadrada) para detectar factores de riesgo transversales.
2. **Criterio de Capacidad Instalada:** La estimación poblacional basada en infraestructura física ( $N=1,890$ ) podría subrepresentar a estudiantes con esquemas 100% virtuales, centrando la validación en el segmento con actividad híbrida/presencial.
3. **Alcance Local:** Los resultados obtenidos solo son representativos para la sede Casco Santo Tomás y no deben extrapolarse directamente a otras sedes de la UNRC, que pueden presentar variaciones en infraestructura, modalidad y perfil del estudiantado.

4. **Generalización Institucional:** La UNRC cuenta con una diversidad de sedes a nivel nacional. Las diferencias territoriales y operativas entre ellas limitan la generalización completa de los hallazgos de una sola unidad académica.
5. **Limitación Temporal:** La muestra refleja una "fotografía" de la percepción estudiantil al cierre de 2025. Los resultados son válidos para este periodo, pero las percepciones y desafíos pueden evolucionar con el tiempo.
6. **Naturaleza de los Datos Autorreportados:** Los datos de la encuesta se basan en el autorreporte de los estudiantes. Si bien esto nos da acceso a sus percepciones (como la intención de abandono), está sujeto a sesgos de memoria y deseabilidad social. Los datos de comportamiento reales (como los del SIE o LMS) son, por naturaleza, más objetivos.
7. **Factores No Observados:** El diseño de la encuesta se enfocó en las variables inferidas por el modelo predictivo y marco teórico. Por lo tanto, excluye otros criterios precisos para capturar factores socioeconómicos o emocionales complejos que también pueden incidir en el riesgo de abandono y que requerirían un instrumento psicométrico más extenso.

Precisamente, esta validación confirma la necesidad de integrar en el sistema SAREP un módulo de 'Encuesta Vocacional' continuo, ya que un sondeo estático es insuficiente para capturar la evolución de estos factores complejos a lo largo del tiempo.

Pese a estas limitaciones inherentes a cualquier estudio de campo, la combinación del muestreo aleatorio simple con una determinación estadística del tamaño de muestra proporciona una **sólida validez interna** al proceso. Los resultados de esta validación local, por lo tanto, no deben interpretarse como una verdad definitiva, sino como lo que son: una **corroboration empírica y contextual** de las hipótesis generadas por el modelo, que sirve como una base metodológica robusta para la futura expansión y refinamiento del sistema SAREP.

### **3.2.4. Resultados y Validación Estadística de Hipótesis Locales**

Con el propósito de realizar una validación empírica local de las relaciones inferidas por el modelo predictivo XGBoost sobre el riesgo de deserción estudiantil, se formularon y contrastaron hipótesis basadas en tres dimensiones centrales: **rendimiento académico, condiciones económico-financieras y expectativas académicas.**

El análisis estadístico de las 100 respuestas arrojó resultados matizados y de gran valor estratégico, permitiéndonos refinar nuestro entendimiento de los factores de riesgo en el contexto específico de la UNRC.

**Tabla-Resumen: Validación Estadística de Hipótesis en la Muestra de la UNRC (n=100)**

Hipótesis	Factor de Riesgo Analizado	Prueba Estadística	p-valor	Resultado y Significado Estratégico
$H_1$	Rendimiento Académico	Chi-Cuadrado ( $\chi^2$ )	0.055	Altamente <b>Tendencial</b> . Confirma que el bajo rendimiento es un factor de riesgo clave, alineado con la "señal de oro" del modelo.
	(Análisis de Intensidad)	Kruskal-Wallis	0.015	<b>Estadísticamente Significativo</b> . Valida que el bajo rendimiento no solo aumenta la probabilidad, sino también la intensidad de los pensamientos de abandono.
$H_{2a}$	Tenencia de Beca	Chi-Cuadrado ( $\chi^2$ )	0.777	<b>No Significativo</b> . No se encontró una asociación estadística directa entre tener beca y la intención de abandono en esta muestra.
$H_{2b}$	Desafío Económico	Chi-Cuadrado ( $\chi^2$ )	0.917	<b>No Significativo</b> . Refuerza la hipótesis del "efecto mediador", donde el estrés financiero impacta

				indirectamente a través del rendimiento.
$H_3$	Expectativas de Carrera (Desajuste Vocacional)	Chi-Cuadrado ( $\chi^2$ )	0.034	<b>Estadísticamente Significativo.</b> Es el predictor directo más fuerte en la muestra local. Subraya la necesidad de intervenciones de orientación vocacional.
	(Análisis de Intensidad)	Kruskal-Wallis	0.016	<b>Estadísticamente Significativo.</b> Confirma que la insatisfacción con la carrera también afecta la intensidad del riesgo.

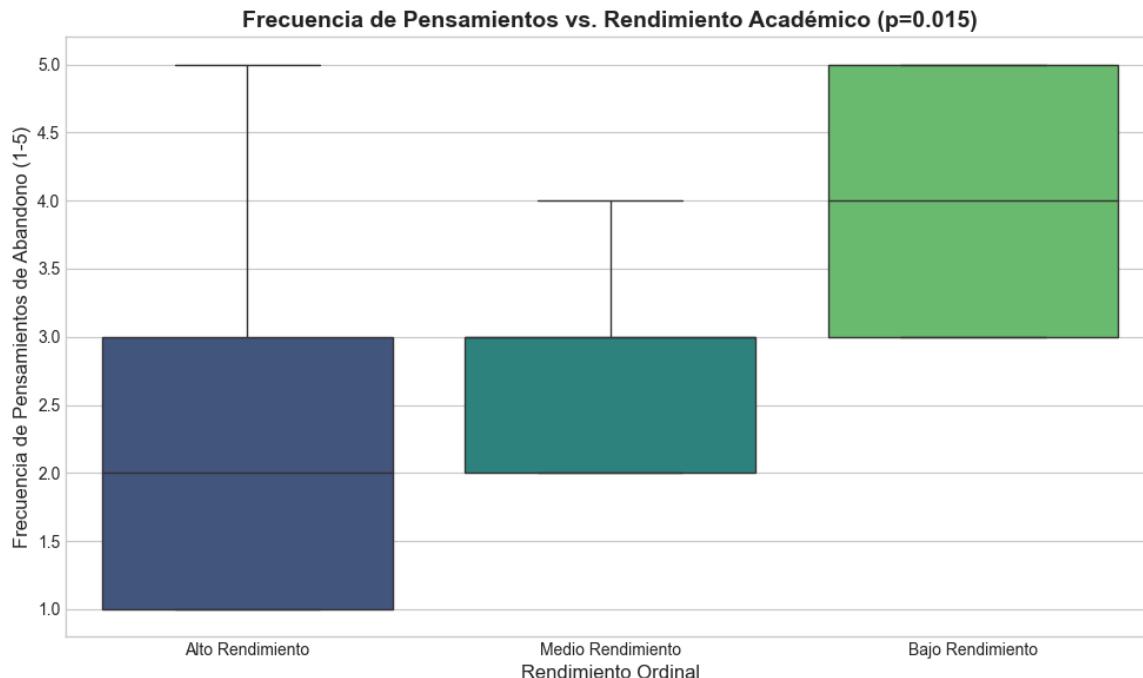
### Hipótesis 1: El Rendimiento Académico (El "Síntoma" Principal)

En línea con la "señal de oro" identificada por el modelo XGBoost, el rendimiento académico mostró una asociación marginal con la decisión de abandonar ( $\chi^2, p = 0.055$ ). Es crucial notar que el valor  $p$  superó ligeramente el umbral estándar de 0.05 debido a una limitación muestral: la baja prevalencia de estudiantes con 'reprobación total' en la sede generó frecuencias esperadas pequeñas, restando sensibilidad a la prueba Chi-Cuadrado.

No obstante, al triangular este resultado con la prueba de **Kruskal-Wallis sobre la intensidad** confirmó una relación claramente significativa ( $p = 0.015$ ) se confirma la hipótesis: aunque pocos alumnos reproban todo, aquellos que lo hacen experimentan una intención de abandono significativamente más intensa.

Esto sugiere que el principal efecto de un bajo rendimiento no es solo "abrir la puerta" a la idea de abandonar, sino intensificar drásticamente su frecuencia, actuando como la "fiebre" que señala una crisis subyacente.

Gráfico de Frecuencia de pensamiento de abando v.s Rendimiento académico



#### **Elaboración propia: Datos de la encuesta**

#### **Hipótesis 2: El "Silencio" de los Factores Financieros y la Hipótesis del Mediador**

Sorprendentemente, nuestro análisis local no encontró una asociación estadística directa entre los factores financieros (beca, dificultades económicas) y la intención de abandono, ni en las pruebas de asociación ni en las de intensidad mostraron una asociación estadística considerable en esta muestra ( $\chi^2, p > 0.7$ ).

Este resultado contraintuitivo no invalida la importancia del estrés económico, sino que propone una dinámica más compleja y refuerza una hipótesis más sofisticada: el efecto mediador:

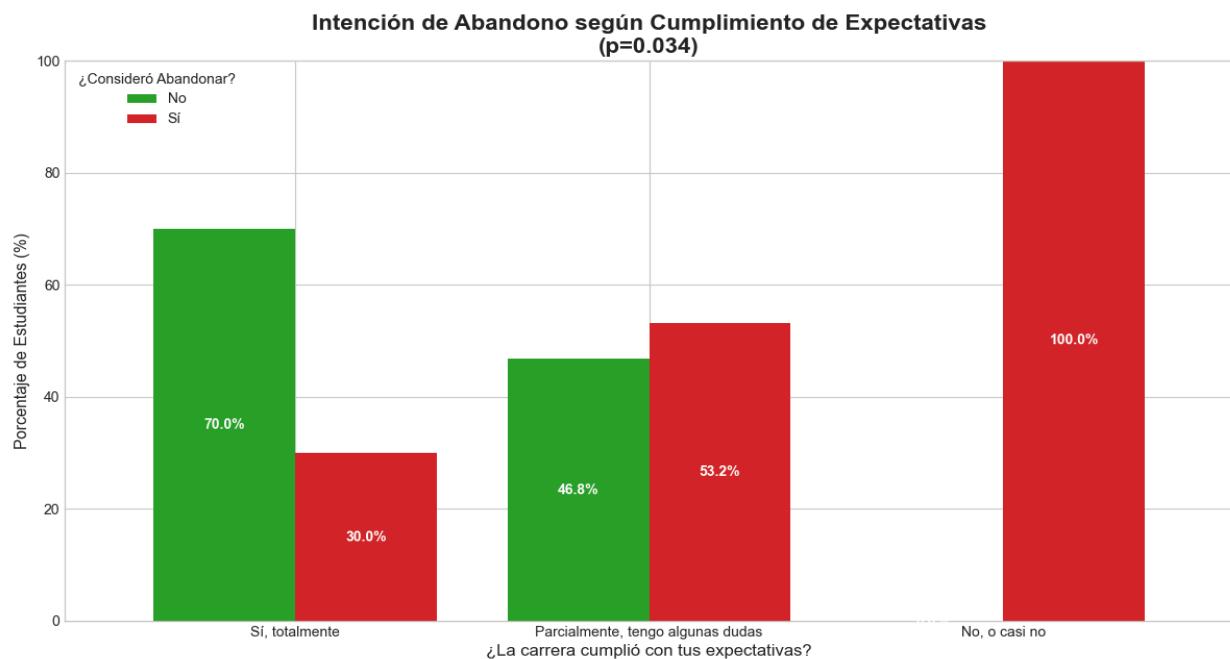
1. **Homogeneidad Relativa:** La población de la UNRC, al ser una universidad pública inclusiva, podría presentar una menor variabilidad socioeconómica, reduciendo el poder estadístico de estas variables.
2. **Sensibilidad del Instrumento:** Las preguntas categóricas de la encuesta pueden no capturar la intensidad real del estrés financiero.
3. **La Hipótesis del Mediador (la más robusta):** Como sugiere el influyente meta-análisis de Robbins et al. (2004), los factores externos y psicosociales a

menudo no impactan directamente la persistencia, sino que lo hacen a través del rendimiento académico. Es plausible que el estrés financiero en la UNRC no cause directamente la intención de abandono, sino que erosione primero el rendimiento, que a su vez se convierte en el disparador principal.

### Hipótesis 3: El Desajuste Vocacional (Estadísticamente Significativo)

La validación local arrojó un hallazgo de particular importancia: el desajuste vocacional es el factor de riesgo con la asociación estadística más fuerte ( $\chi^2, p = 0.034$ ) con la intención de abandono. Este descubrimiento, alineado con la importancia de la variable Course en nuestro modelo predictivo, subraya la necesidad crítica de que SAREP integre mecanismos de orientación vocacional temprana.

La evidencia visual es contundente: el 100% de los estudiantes que reportaron estar 'Insatisfechos' con su carrera también consideraron abandonar, en marcado contraste con solo el 29% de los que se sienten 'Satisfechos'. Adicionalmente, el análisis kruskal-Wallis de intensidad reveló una significancia ( $p = 0.016$ ).



**Elaboración propia: Datos de la encuesta**

### **Conclusión Metodológica General: Del Síntoma a la Mentalidad**

1. La estrategia dual de la investigación —un prototipo predictivo para el descubrimiento y una validación local para la corroboración— nos lleva a una conclusión final que resuena con la investigación de vanguardia de Claro, Paunesku y Dweck (2016) sobre la "mentalidad de crecimiento".
2. Nuestros hallazgos, tanto del modelo como de la encuesta, demuestran que, si bien no podemos ignorar las barreras estructurales (como las socioeconómicas), los factores psicológicos (como el desajuste vocacional) y sus manifestaciones conductuales (el rendimiento académico) son los predictores más directos y accionables.
3. Esto refuta un determinismo simplista y valida la **arquitectura de datos propuesta**: el sistema no debe limitarse a importar calificaciones; debe incluir sensores activos para medir el desajuste vocacional. Al monitorear el rendimiento académico como 'síntoma' y la vocación como 'causa', SAREP se posiciona no solo como un tablero de control, sino como una herramienta de diagnóstico preciso.

### **3.3. Fase 3: Benchmarking de Soluciones para la Identificación de “Palancas de Acción”**

Finalmente, para asegurar que la solución propuesta estuviera fundamentada en las mejores prácticas globales, se realizó un estudio de benchmarking exhaustivo. Este análisis arrojó lecciones estratégicas cruciales de referentes nacionales (UNAM, IPN, UAM) e internacionales (GSU, ODU, EAB), garantizando que la propuesta de SAREP sea creíble, viable y esté fundamentada en evidencia rigurosa.

## **4. Solución Propuesta: El Sistema SAREP**

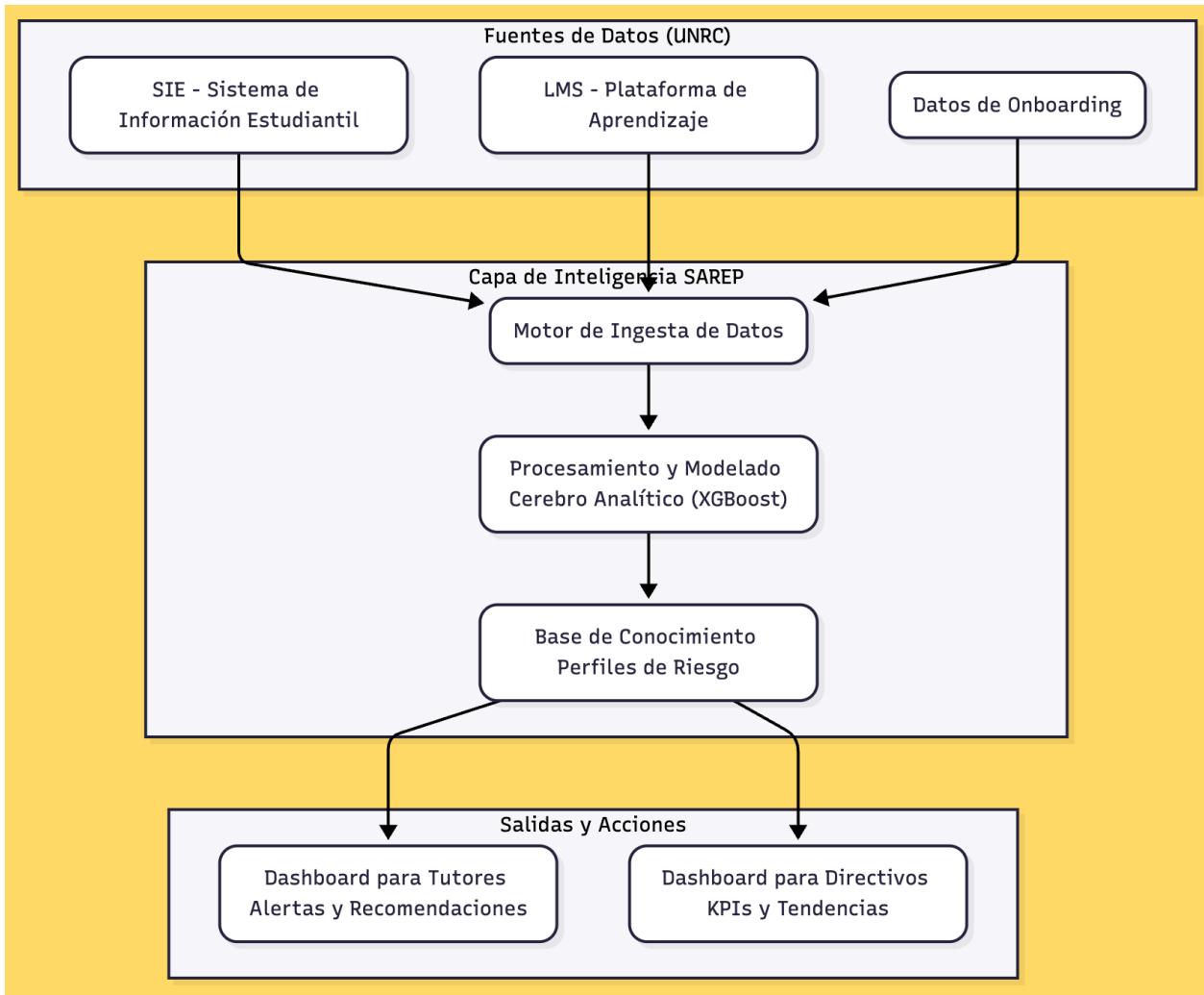
Con base en el diagnóstico del problema y el análisis de las mejores prácticas, se propone la implementación del **Sistema de Acompañamiento y Retención Estudiantil Proactivo (SAREP)**. Esta sección detalla la arquitectura de la solución, desde su visión conceptual hasta los componentes técnicos prototipados, demostrando su coherencia, viabilidad y poder transformador.

#### **4.1. Visión y Arquitectura General: El "Waze" del Viaje Académico**

La visión de SAREP se articula a través de una analogía poderosa: ser el "Waze" del viaje académico del estudiante. Al igual que Waze no construye nuevas carreteras, SAREP no reemplaza la infraestructura existente de la UNRC (LMS, SIE), sino que actúa como una **capa de inteligencia** que se superpone a ella. Su función es unificar datos, analizar información para predecir obstáculos (riesgo de abandono) y proporcionar rutas de apoyo proactivo para asegurar que cada estudiante llegue a su destino: la graduación.

La implementación de esta visión requiere una arquitectura de datos robusta que gestione el flujo de información de extremo a extremo.

#### **Exhibit 5: Arquitectura General del Pipeline de Datos ELT de SAREP**



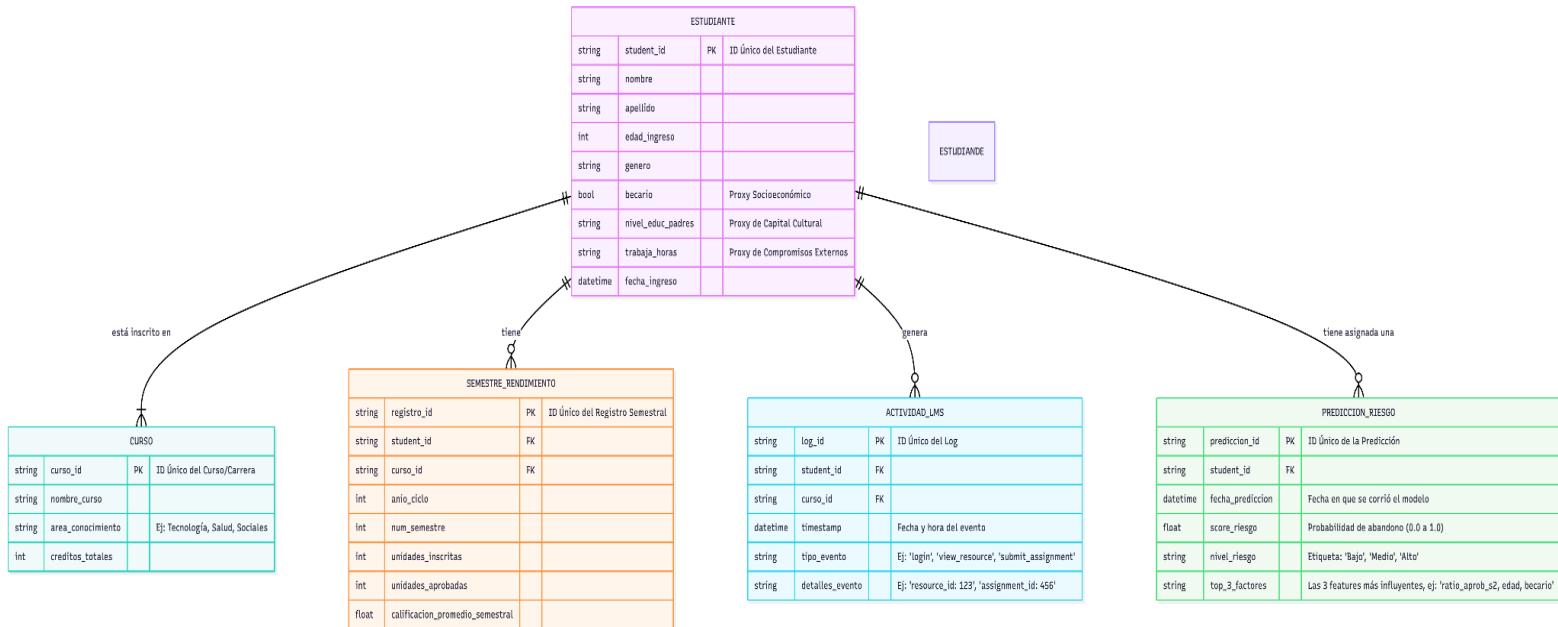
Fuente: Elaboración propia

## 4.2. Diseño de la Base de Conocimiento: Un Modelo de Datos para un Perfil Estudiantil de 360 Grados

La efectividad de SAREP reside en su capacidad para estructurar la información de manera relacional. Un modelo predictivo es tan bueno como los datos que lo alimentan, y una base de datos bien diseñada es la columna vertebral que garantiza la calidad, integridad y accesibilidad de esa información.

Para esto se ha diseñado un Esquema Entidad-Relación (ERD) normalizado que modela la base de conocimiento del sistema. Este diseño va más allá de una simple tabla, creando un ecosistema de datos interconectados que permite un análisis complejo y longitudinal.

## Esquema Entidad-Relación (ERD) de la Base de Conocimiento de SAREP



Fuente: Elaboración propia

## 4.4. Componente 2: El Ecosistema de Intervención Proactiva y el Dashboard del Tutor

La predicción algorítmica es solo el primer paso; su valor se materializa únicamente cuando se traduce en una acción humana informada. Por ello, el segundo componente de SAREP es un **Ecosistema de Intervención Proactiva**, cuyo corazón operativo es el **Dashboard del Tutor**. Esta herramienta está diseñada para cerrar el ciclo "de análisis a acción" y convertir las predicciones de riesgo en estrategias de apoyo concretas y personalizadas.

## *Círculo de Intervención Proactiva*



### *El Dashboard del Tutor: De la Alerta a la Acción*

Para ilustrar cómo funcionaría este ecosistema, se desarrolló un prototipo funcional del Dashboard del Tutor utilizando Streamlit. Este "Simulador de Riesgo para la UNRC" no es una aplicación de producción, sino una demostración conceptual que visualiza cómo los hallazgos de nuestro análisis se convierten en una herramienta accionable.

La interfaz presentaría a cada tutor una lista de sus estudiantes, priorizada no alfabéticamente, sino por el nivel de riesgo calculado por el "Cerebro Analítico". Al seleccionar a un estudiante, el tutor vería un perfil detallado.

## Mockup del Dashboard del Tutor SAREP (Simulación UNRC)



### *La Lógica Interna: Un Motor de Reglas Basado en Evidencia*

El prototipo del dashboard está impulsado por una lógica que simula cómo los datos específicos de la UNRC se traducirían en una predicción de riesgo. En lugar de un mapeo simple, utiliza un "motor de reglas de negocio" que calcula un Puntaje de Riesgo Contextual.

Este puntaje pondera los factores basándose directamente en la evidencia estadística recopilada en nuestra encuesta de validación local (la prueba Chi-Cuadrado). Por ejemplo:

- El **desajuste vocacional** (nuestro hallazgo más significativo,  $p=0.028$ ) recibe el **mayor peso** en el cálculo del riesgo.
- El **rendimiento académico** (altamente tendencial,  $p=0.052$ ) recibe un peso considerable.

- Los **factores socioeconómicos** (no significativos en el análisis local) actúan como "**multiplicadores**" de riesgo: su impacto se considera principalmente cuando ya existen otros problemas académicos o vocacionales, modelando así el **efecto mediador** que identificamos.

Este enfoque asegura que las alertas generadas no solo sean técnicamente precisas (basadas en el motor XGBoost), sino también contextualmente relevantes para la realidad de la población estudiantil de la UNRC. El resultado es un sistema que le dice al tutor no solo *quién* está en riesgo, sino, con alta probabilidad, *por qué*, permitiendo una intervención verdaderamente personalizada.

## 5. Análisis de Viabilidad y Caso de Negocio

### 5.1. Impacto Probado: Evidencia Cuantitativa

La propuesta de SAREP no es un experimento. Se basa en modelos de éxito probados: Georgia State University (GSU): Su ecosistema “construido” logró un aumento de 23 puntos porcentuales en su tasa de graduación.

EAB Navigate360: Plataformas comerciales “compradas” que reportan un ROI típico de 5:1.

### 5.2. Análisis de Costos: Construir vs. Comprar

Al contrastar el mercado de software educativo internacional con los costos laborales del sector tecnológico en México, el análisis favorece decisivamente al desarrollo interno (“In-House”).

#### Comparativa de Costos Proyectada (Anual):

- **Opción A: Licenciamiento Comercial (SaaS)**
  - Basado en plataformas estándar (ej. Salesforce Education / EAB) para 57,000 estudiantes.
  - Costo estimado: ~\$5 USD/estudiante/año.
  - **Total Anual Recurrente: ~\$285,000 USD (Costo perpetuo).**

- *Desventaja:* Dependencia del proveedor y costos atados al tipo de cambio.
- 
- **Opción B: Desarrollo SAREP In-House (Recomendada)**
  - Equipo dedicado ("Tiger Team") en CDMX: 1 Lead Data Scientist, 1 Data Engineer, 1 Full Stack Developer.
  - Infraestructura Cloud (AWS/Azure): Escalable según uso.
  - **Total Estimado Año 1 (Desarrollo): ~\$150,000 USD.**
  - **Total Estimado Año 2+ (Mantenimiento): ~\$80,000 USD.**

Veredicto: La construcción propia genera un ahorro del 47% en el primer año y de más del 70% en años subsecuentes. Además, garantiza la Soberanía de Datos de la UNRC, permitiendo adaptar los algoritmos a la realidad socioeconómica específica de sus estudiantes sin depender de "cajas negras" extranjeras.

**Tabla 1: Comparativa de Costos y Beneficios (Proyección Anual)**

Rubro	Solución Comercial (SaaS)	Solución In-House (SAREP)
Proveedor Tipo	EAB, Salesforce, Ellucian	Equipo Interno (Tiger Team)
Modelo de Costo	Licencia por Estudiante + Implementación	Nómina (Salarios CDMX + Infraestructura Nube)
Costo Implementación (Año 1)	~\$100,000 USD (Consultoría y Configuración)	Incluido en Nómina Operativa
Costo Recurrente (Anual)	~ 285,000USD( 285,000USD( 5 USD x 57k alumnos)	~\$135,000 USD (Nómina + Nube AWS/Azure)
Costo Total Año 1	~\$385,000 USD	~\$135,000 USD
Ahorro Estimado	–	~65% vs. Comercial

<b>Tiempo de Despliegue</b>	6-9 Meses	6-8 Meses (MVP)
<b>Propiedad de Datos</b>	Proveedor (Vendor Lock-in)	Total (Soberanía UNRC)
<b>Flexibilidad</b>	Baja (Depende del Roadmap del proveedor)	Total (Adaptable a Desajuste Vocacional)

#### **Análisis de TCO (Costo Total de Propiedad):**

Mientras que la solución comercial impone un costo perpetuo atado al dólar y al volumen de estudiantes (a más éxito de la universidad, más paga), la solución In-House tiene un costo marginal decreciente. Una vez construido el núcleo del sistema (Año 1), el mantenimiento requiere menos recursos, y el costo de añadir nuevos estudiantes a la base de datos es de centavos en infraestructura de nube, no de dólares en licencias. Esto alinea financieramente el proyecto con la misión de austeridad y eficiencia de la universidad pública.

## **6. Recomendaciones Estratégicas y Hoja de Ruta**

### **6.1. Hoja de Ruta Física para la Implementación**

- **Fase 1:** Cimentación y Calibración (Meses 1-4)
  - Ingesta de Datos Reales: Conexión del Pipeline ELT diseñado (Sección 4.1) con el Sistema de Control Escolar (SIE) histórico de la UNRC.
  - Re-Entrenamiento del Modelo: Sustitución del dataset proxy por los datos históricos de la UNRC para recalibrar los pesos del algoritmo XGBoost a la realidad local.
- **Fase 2:** Piloto Controlado - "Sede Casco Santo Tomás" (Meses 5-8)
  - Despliegue del Dashboard en una sola unidad académica.
  - Integración del Módulo Vocacional: Implementación de micro-encuestas digitales para detectar el "Desajuste Vocacional" (Hipótesis 3).

- Capacitación de un grupo selecto de tutores y recepción de feedback.
- **Fase 3:** Despliegue Institucional (Meses 9-12)
  - Escalado de la infraestructura en la nube para soportar los 57,000 estudiantes.
  - Automatización total de alertas de riesgo.

## 6.2. Recomendación de Política Institucional

Se recomienda establecer a SAREP no como un proyecto de TI, sino como un activo de Investigación Institucional. La UNRC debe capitalizar su propia data para crear inteligencia, en lugar de subcontratar el entendimiento de sus estudiantes.

Para que las soluciones basadas en inteligencia sean efectivas, la UNRC debe transitar de un enfoque de "registro administrativo" a uno de "inteligencia activa". Esto implica dos mandatos normativos:

1. **Estandarización de la Interacción Digital:** Se debe normar el uso de plataformas educativas. Si bien se permite flexibilidad pedagógica, se debe establecer la obligatoriedad de registrar hitos de evaluación parciales (no solo finales) en sistemas centralizados. Sin datos intermedios, no hay alerta temprana posible.
2. **Privacidad por Diseño:** La centralización de datos sensibles (socioeconómicos y académicos) debe realizarse bajo un protocolo estricto de Gobernanza de Datos. Se recomienda la implementación de "Murallas Chinas" digitales (RBAC) que permitan al equipo de Ciencia de Datos entrenar modelos con información pseudonimizada, cumpliendo con la normativa de protección de datos, mientras que solo los tutores autorizados acceden a la identidad del estudiante en riesgo.

## 6.3. Recomendación Táctica: Estabilización e Ingesta Flexible

Se ha identificado que la adopción irregular del LMS institucional (Moodle) obedece tanto a factores culturales como a desafíos de infraestructura técnica (inestabilidad o

latencia). Forzar una migración total inmediata sin resolver la estabilidad técnica pone en riesgo la continuidad de los datos. Por tanto, se recomienda una estrategia de dos carriles:

1. **Carril de Estabilización (IT):** Solicitar una auditoría de infraestructura al área de Tecnología para diagnosticar y resolver las caídas del LMS, con el objetivo de establecerlo como la fuente única de la verdad en un plazo de 12 a 18 meses.
2. **Carril de Interoperabilidad (SAREP - Fase 1):** Para no detener el proyecto mientras se estabiliza el LMS, se deben desarrollar mecanismos de ingesta resilientes. Esto implica habilitar la carga de archivos planos (CSV/Excel) estandarizados para cortes parciales y explorar conectores API con plataformas de contingencia (como Google Classroom).

**Objetivo:** Asegurar que el modelo predictivo reciba datos de rendimiento hoy, independientemente de la plataforma que el docente logre utilizar, mientras se trabaja en paralelo hacia la estandarización institucional.

## 7. Conclusión

La Universidad Nacional Rosario Castellanos se encuentra ante una encrucijada histórica: mantener un modelo de "puerta giratoria", donde el éxito en la matrícula se diluye por la ineficiencia terminal, o transitar hacia una "vía de graduación" impulsada por inteligencia institucional.

Este análisis demuestra que la barrera principal no es la falta de capacidad del estudiante, sino la ausencia de mecanismos de detección temprana. Actualmente, la información vital sobre el rendimiento se pierde en silos desconectados o registros tardíos, haciendo imposible la intervención preventiva.

La solución propuesta, el sistema SAREP, valida que la respuesta no reside en adquirir costosas licencias de software comercial, sino en desarrollar soberanía tecnológica. La opción de desarrollo In-House no solo representa un ahorro superior al 60% frente al mercado, sino que permite lo que ningún proveedor externo puede ofrecer: integrar la

realidad fragmentada de las aulas (desde plataformas oficiales hasta herramientas auxiliares) en un cerebro analítico, unificado y seguro.

En última instancia, implementar SAREP es una decisión de justicia social. Significa dejar de administrar calificaciones pasadas para empezar a gestionar futuros posibles, asegurando que la promesa de movilidad social de la UNRC se cumpla para cada uno de sus 57,000 estudiantes.

## 8. Referencias

### Fuentes Primarias Institucionales y Gubernamentales

1. Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES). (s.f.). Anuario Estadístico de Educación Superior. Recuperado de <https://www.anuies.mx/informacion-y-servicios/informacion-estadistica-de-educacion-superior/anuario-estadistico-de-educacion-superior>
2. Centro de Estudios Espinosa Yglesias (CEEY). (s.f.). Movilidad Social. Recuperado de <https://ceey.org.mx/movilidad-social/>
3. Gobierno de México. (2025). Primer Informe de Gobierno 2025. Presidencia de la República.
4. Instituto Mexicano para la Competitividad, A.C. (IMCO). (2024). Compara Carreras 2024: ¿Formamos el talento que México necesita? Recuperado de <https://imco.org.mx/compara-carreras-2024-formamos-el-talento-que-mexico-necesita/>
5. Instituto Mexicano para la Competitividad. (2024, agosto). Compara Carreras 2024: Las carreras mejor pagadas en México [Comunicado de prensa]. IMCO. [https://imco.org.mx/wp-content/uploads/2024/08/Comunicado\\_ComparaCarreras-2024.pdf](https://imco.org.mx/wp-content/uploads/2024/08/Comunicado_ComparaCarreras-2024.pdf)
6. Instituto Nacional de Estadística y Geografía (INEGI). (2024). Tasa de abandono escolar por entidad federativa según nivel educativo, ciclos escolares

seleccionados de 2000/2001 a 2023/2024. Recuperado de  
[https://www.inegi.org.mx/app/tabulados/interactivos/?pxq=Educacion\\_Educacion\\_11\\_c6aa7c65-4d89-4eaf-972e-431727fc686d](https://www.inegi.org.mx/app/tabulados/interactivos/?pxq=Educacion_Educacion_11_c6aa7c65-4d89-4eaf-972e-431727fc686d)

7. Instituto Politécnico Nacional (IPN). (2024). Agenda Estadística 2024. Recuperado de  
[https://www.ipn.mx/assets/files/coplaneval/docs/dii/agenda\\_estadistica/agenda-e\\_stadistica-2024-f.pdf](https://www.ipn.mx/assets/files/coplaneval/docs/dii/agenda_estadistica/agenda-e_stadistica-2024-f.pdf)
8. Universidad Autónoma Metropolitana (UAM). (2024). Informe de actividades UAM 2024. Docencia. Recuperado de  
[https://transparencia.uam.mx/inforganos/rg/2024/rep/informe-uam-2024\\_docencia.pdf](https://transparencia.uam.mx/inforganos/rg/2024/rep/informe-uam-2024_docencia.pdf)
9. Universidad Nacional Rosario Castellanos. (2023). Numeralia 4o Trimestre 2023. Portal de Datos Abiertos. Recuperado de  
<https://rcastellanos.cdmx.gob.mx/numeralia-2023/numeralia-4o-trimestre-2023/maatricula>
10. Universidad Nacional Rosario Castellanos. (2024). Numeralia 4o Trimestre 2024. Portal de Datos Abiertos. Recuperado de  
<https://rcastellanos.cdmx.gob.mx/numeralia-2024/numeralia-4o-trimestre-2024>

### Fuentes de Benchmarking y Casos de Éxito

11. EAB. (s.f.). Navigate360 Resource Hub. Recuperado de  
<https://eab.com/partner-hub/navigate360-resource-hub/>
12. Georgia State University. (s.f.). Student Success Initiatives. Recuperado de  
<https://success.gsu.edu/>
13. Ithaka S+R. (2016). Building a Pathway to Student Success at Georgia State University. Recuperado de

<https://sr.ithaka.org/publications/building-a-pathway-to-student-success-at-georgia-state-university/>

14. Predict students' dropout and academic success. (s.f.). UCI Machine Learning Repository. Recuperado de <https://archive.ics.uci.edu/dataset/602/predict+students+dropout+and+academic+success>

### Fuentes Académicas y Metodológicas Clave

15. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 177-194.
16. Castillero, O. et al. (2019). Descripción del abandono escolar y su costo económico en la... Revista Janaskakua. Recuperado de <https://publicaciones.umich.mx/revistas/janaskakua/enfermeria/article/view/93>
17. Rodríguez, R. (2022). Estudiantes “no tradicionales” en la educación superior. *Revista de la Educación Superior*.
18. Georgia State University. (2019). \*Case Study: Georgia State University\*. AGB. Recuperado de [https://agb.org/wp-content/uploads/2019/01/case\\_study\\_innovation\\_georgia.pdf](https://agb.org/wp-content/uploads/2019/01/case_study_innovation_georgia.pdf)
19. Purdue University. (2010). Signals: Applying Academic Analytics. \*EDUCAUSE Review\*. Recuperado de <https://er.educause.edu/articles/2010/3/signals-applying-academic-analytics>
20. Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. \*Psychological Bulletin\*, 130\*(2), 261–288. <https://doi.org/10.1037/0033-2909.130.2.261>

21. Sonrai Analytics. (2025). \*What is XGBoost? - and how can it be used in Precision Medicine?\* Recuperado el 26 de octubre de 2025, de <https://sonraianalytics.com/what-is-xgboost/>
22. Tinto, V. (1993). \*Leaving college: Rethinking the causes and cures of student attrition\* (2nd ed.). The University of Chicago Press
23. Claro, S., Paunesku, D., & Dweck, C. S. (2016). Growth mindset tempers the effects of poverty on academic achievement. *Proceedings of the National Academy of Sciences*, 113(31), 8664–8668.  
<https://doi.org/10.1073/pnas.1608207113>
24. Universidad Nacional Rosario Castellanos. (2024). Informe de Resultados 2019–2024 (Versión B). Gobierno de la Ciudad de México. Recuperado de [https://www.rcastellanos.cdmx.gob.mx/storage/app/media/URC%20-%20Informe%20Resultados%202019-2024%20B\\_.pdf](https://www.rcastellanos.cdmx.gob.mx/storage/app/media/URC%20-%20Informe%20Resultados%202019-2024%20B_.pdf)
25. Secretaría de Economía. (s.f.). Profesionistas y técnicos. Data México. Recuperado el 4 de noviembre de 2025, de <https://www.economia.gob.mx/datamexico/es/profile/occupation/profesionistas-y-tecnicos>